COMMUNITY ANNOTATIONS USER MANUAL

This manual is intended to explain in detail how researchers can enter the MGOS database and in particular, how to add manual annotation information to the *Magnaporthe* genome. We hope it will stimulate researchers to add their sequence, expression and mutant data to provide maximum utility of the information to the community.

Table of Contents

(A) General Information about Database:

(1) Information
(2) Forum
(3) 4th IRBC abstracts
(4) Gene
(5) EST-contig
(6) RL-SAGE
(7) Microarray
(8) Mutants
(9) Genome
10) BLAST
11) BLAT
12) Literature
13) Links

(B) Step by step guide to add annotation information:

(1) Registration
(2) Login
(3) Forgot your Password
(4) Annotation of a particular locus
(5) Entry on Gene page
(6) To add information
(7) Sections of Gene page
    (A) Gene Symbol
    (B) Synonyms
    (C) Gene Description
    (D) Gene Comments
    (E) Secreted protein
    (F) Gene Ontology Terms
    (G) Fungal Anatomy Terms
    (H) Publications
      (i) Add a citation
      (ii) Edit a citation
    (I) Mutants
      (i) Add a Mutant
      (ii) Edit a Mutant
    (J) Transcript annotation

(A) General Information about Database:

For access to the community annotation feature of the *M. oryzae* genome, one goes to the MGOS database home page (http:// www.mgosdb.org/). MGOS is divided into several subheadings:

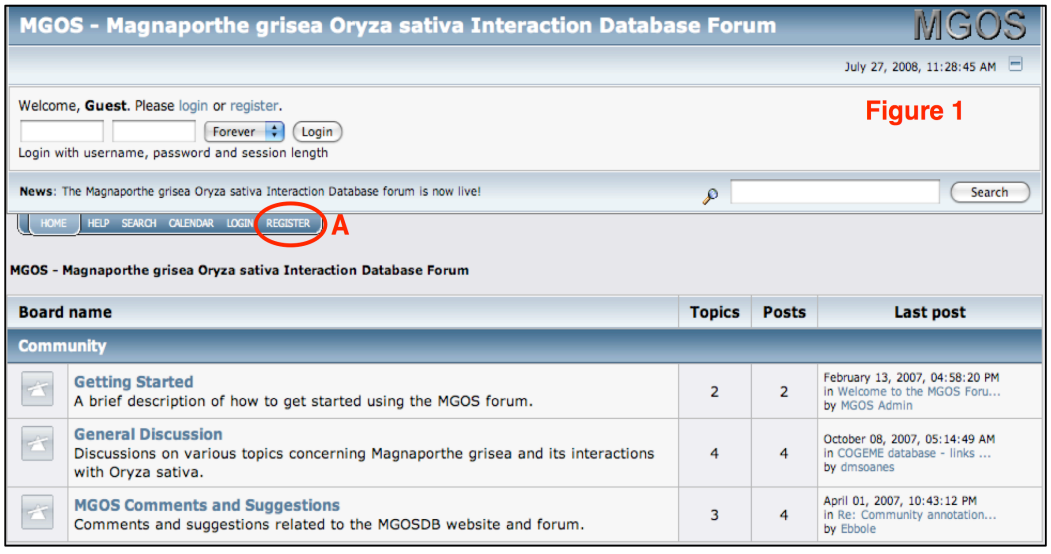
The First section includes Information of general interest about the MGOS project and to the Magnaporthe community.

1) **Information**: This contains a basic description of the history and researchers involved in generating the database.

2) **Forum**: This page allows one to register as a community annotator. The steps are as follows:
   a) Click on forum.
   b) You will see several options as shown in **Figure 1**.
   c) Click on register (**Figure 1A**).
   d) This will take you to the page where you can enter your e-mail address for MGOS.
   e) After registering you can use that e-mail address to get into the forum and you can use several available features there. These are:
      i) Join or start a discussion thread on a topic of community interest
      ii) Leave comments or suggestions for using MGOS, features you would like to see added, comments on existing data, etc.

3) **4th IRBC Abstracts**: access to the abstracts of the Foruth International Rice Blast Conference (Changsha, China October 9-14, 2007).

Figure 1



*The Second section of the database contains the **DATA**, divided into several sub-sections.*

4) **Gene:** Following this link opens up gene search page, which is the primary page for selecting genes for manual annotation, and for viewing features of genes of interest. On this page you can search for individual *Magnaporthe oryzae* locus*,* either by entering an MGOS number (MC_**xxxxx)** or the associated Broad number (**MGGxxxxx**). Moreover a **locus name can be written** in the Basic Gene Search window.

5) **EST Contigs:** Allows searching of ESTs from libraries prepared under several different growth conditions. All ESTs in Genbank are available here.

6) **RL-SAGE:** This section provides access to SAGE data developed in the laboratory of Dr. Guo-Liang Wang at Ohio State.

7) **Microarray:** Under development

8) **Mutants:** To access information on *Magnaporthe* mutants generated by NSF plant Genome grant, select this link.
9) **Genome:** To access the Magnaporthe genome browser, select this link.

The Third section of the database contains search tools to help the community to find and annotate loci of interest, access to the Magnaporthe literature, and links of interest.

10) **BLAST**: This link is used to search for a particular locus of interest in either MGOS, or Uniprot databases, by entering either a DNA or protein sequence.

11) **BLAT:** BLAT is used to find sequences of high identity and of a minimum length and works faster than BLAST. For DNA alignments, the sequences must be greater than 95% and at least 40 nt in length and for protein alignments, greater than 80% identical over 20 amino acids.

12) **Literature:** The available literature in MGOS can be searched using this option.

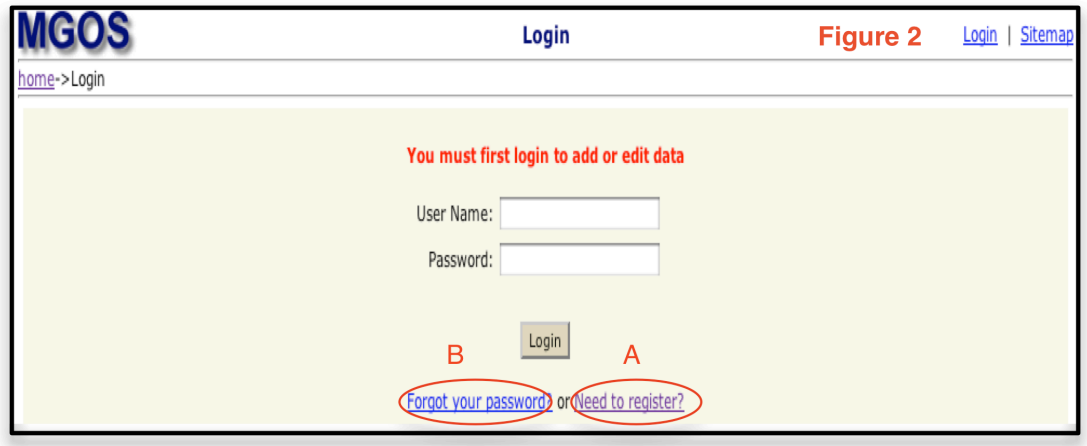
13) **Links:** Additionally useful database links along with other additional resource links are available under this option.

(B) Step by step guide to add annotation information:

There are several steps to annotate a locus:

1) **Registration**:
(a) To add annotation information to MGOS you need to register as described above (**Figure 1**) or
(b) By clicking on the main webpage of MGOS, there is a link for *M. oryzae* community annotation in front of the gene column. Clicking on this link will open the page shown in **Figure 2**. Follow **Figure 2A** for direct access to the registration page

Figure 2



for MGOS. If you have any questions, feel free to contact Anupreet kour (akjosan@email.arizona.edu) or Kevin Greer (kevin@agcol.arizona.edu). Full contact information is available at the end of this manual.

2) **Login:**
(Figure 2) To Login enter your User Name and Password and click "Login".

3) **If you forget your password**: you need to click on "Forget your password?" (**Figure 2B**). You will be sent an email to enter a new password. Or you can contact MGOS administration for more information.

(4) **For annotation of a particular locus:**
Clicking on the *M.oryzae* community annotation link on the front page will take you to the annotation page as shown in **Figure 2**. After Login on this page as described above, you will reach the page as shown in **Figure 3**.
Either you perform a BLAST search with nucleotide and amino acid sequence to identify the locus of interest (**Figure 3B**) or use the Broad or MGOS locus number to search the gene of interest (**Figure 3A**). All the instructions for annotation are given on this link.

(5) **Next, enter the "Gene" page by clicking on the locus of interest.** It will take you to the gene edit page as shown in Figure 4**.** Show audits is the option, that will take you to page showing about the annotations regarding that gene which includes curator's name and date of curation.

(6) **To add information**: The gene edit page (**Figure 4**, below) is the link, where you begin data entry. All fields in "green" are editable. All editable fields will also have either a pencil icon or a green underlined link for entry. The original information is shown in black. For editing, click on the icon or link, add information and click the Save icon (diskette) to the left of the box. If you incorrectly opened a field, click on the cancel icon.

Figure 3

MGOS
Community Annotations
Figure 3
Login | Sitemap
home->Community Annotations
Search for Gene to Annotate:
A
Magnaporthe oryzae:
Search
(eg. MC_15370, MGG_15370, AVR-Pita, CUT1, 40S ribosomal)
Blst Sequence:
Paste your nucleotide or amino acid query sequence in field below
BLAST
B
Instructions:
1. To begin annotating a gene you may either search based on its name, symbol, or accession number using the basic gene search below, or you may blast the sequence against the existing gene models by pasting your sequence into the text box in the blast section below and clicking the BLAST button.
2. Depending on the results of your search you may be presented with a list genes, at which point you simply need to click on the gene accession number (which begins with MC_) that you are interested in to begin the annotation process. In other cases you will be taken directly to the annotation screen.
3. If your gene cannot be located using either on of these methods please contact us at mgos@mgosdb.org so that we can provide further assistance.
4. You may also perform a more elaborate serch for genes to annotate using the Advanced Gene Search page or provide additional parameters to the blast algorith using the Blast page.

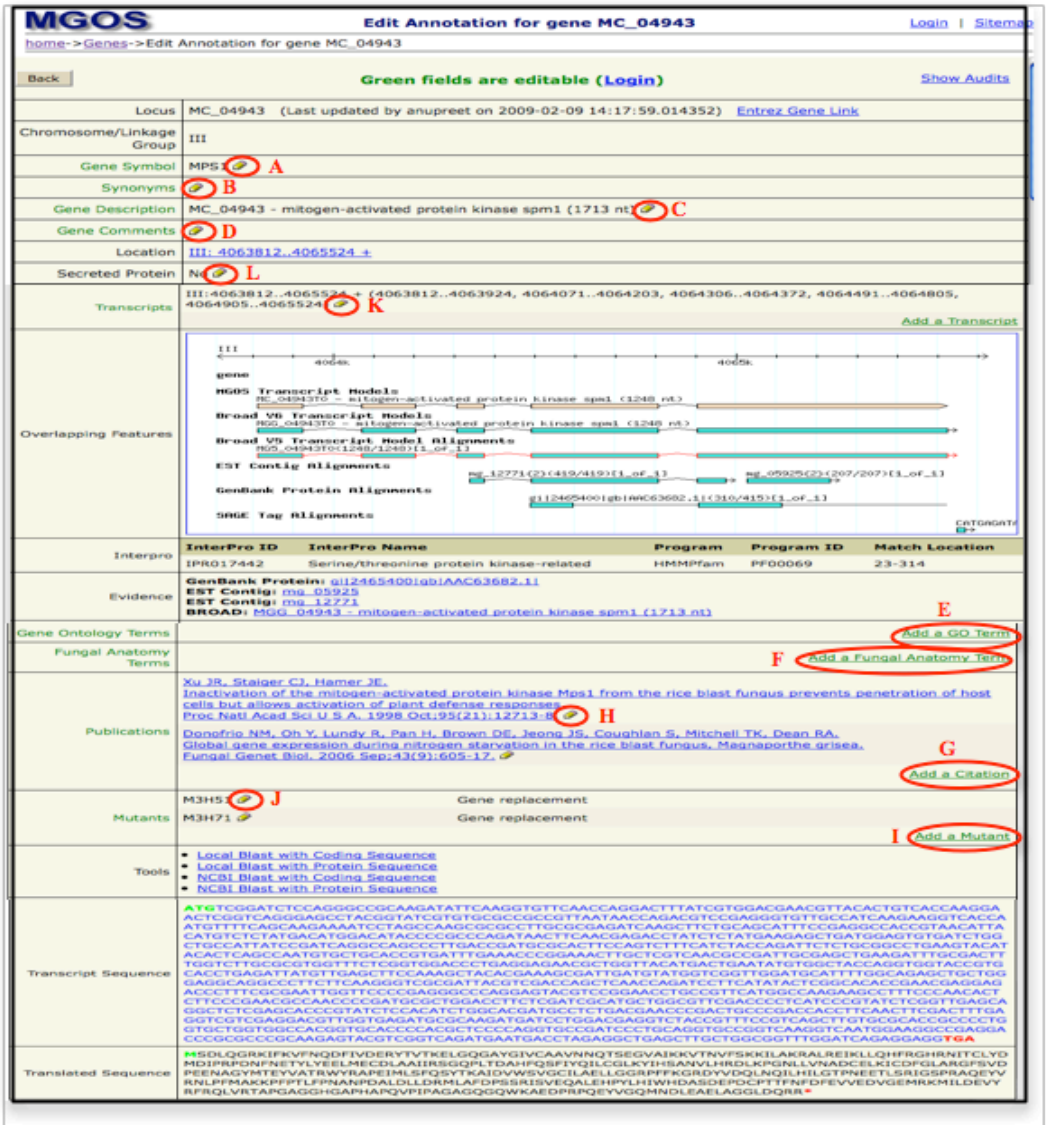
(7) **Sections of the Gene Page:** Gene page is further divided into several sections that can be annotated.

(A) **Gene symbol (Figure 4A)**: To add or correct a gene symbol, click on the pencil icon insert your information and save it (**Figure 5**). *M. grisea* gene symbols are usually three letters and a number (e.g., *MFP1, HDL1*, *CYP1* etc.). Please see the article by Yoder et al for guidelines on proper *M. grisea* gene nomenclature (Yoder, O. C., B. Valent, and F. Chumley. 1986. Genetic nomenclature and practice for plant pathogenic fungi. 76:383-385).

(B) **Synonyms (Figure 4B):** Alternative names for genes go in this field (i.e., mutants identified by different research groups have led to different names for the same gene). For example, *NTH1* is a synonym for PTH9 (MGOS number *MC_09471)* and is inserted into the **Synonyms** field. To add a synonym, select the pencil icon to open the field (**Figure 4B**), insert the synonym, and click on the diskette icon to save. To add more than one synonym, separated each by a space or slash.

(C) **Gene Description (Figure 4C):** In this field you can add a description of the locus if the function of the gene is known. For example you could insert the full gene name like "pathogenicity locus 9" for *PTH9,* for *HDL1* you could enter that it encodes a "hormone sensitive lipase". Add information as described above.

Figure 4

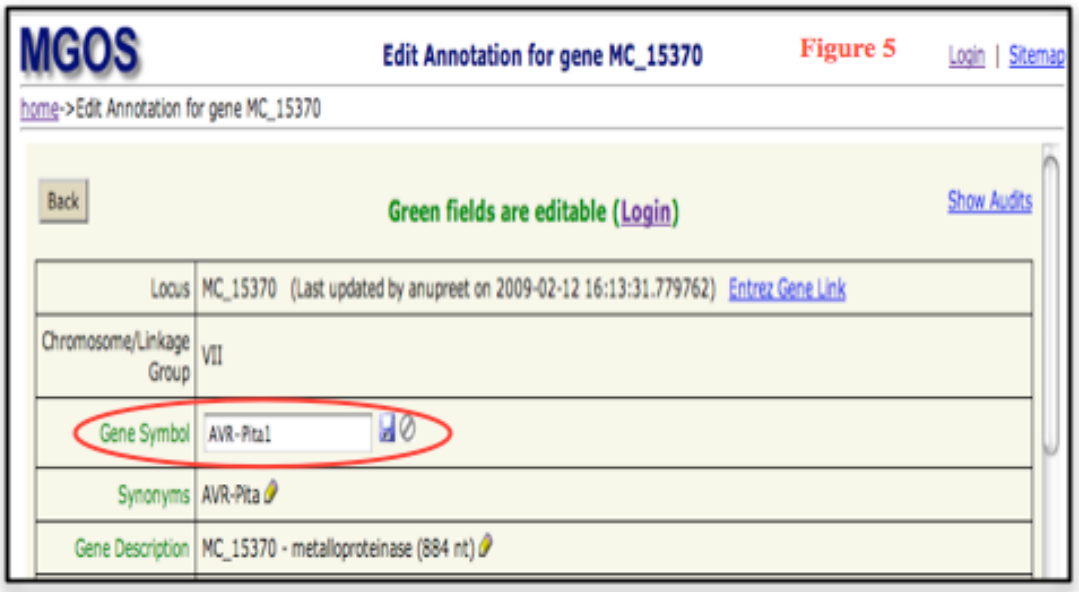


(D) **Gene Comments (Figure 4D):** In this field you may insert short comments about the gene, for example information about its regulation or localization (eg. upregulated during appressorium development, nuclear-localized, etc). Add information as described in Step 5 above.

(E) **Secreted protein (Figure 4L):** The secreted protein information is generated here by automation using reference "e-fungi: a data resource for comparative analysis of fungal genome. Cornelia *etal*. BMC genomics, 2007, 8:426.

(F) **Add a GO term (Figure 4E):** Gene Ontology (GO) terms for biological process, molecular function, and cellular component are added here along with a reference(s). When you click the "**Add a GO term"** link (**Figure 4E**), you will be redirected to a

Figure 5



new page headed **Add Gene Ontology Term for MC_XXXX (the name of the locus)** (Figure 6). Click on **Choose Term** (Figure 6A) to open a search page, which is called **Find Gene Ontology Term** (**Figure 7**).

If you already know the GO ID number, you can search using it. Another way to search is to put your query in quotes (e.g., "cholinesterase activity" or "DNA damage"). If you just put these phrases in without the quotation marks the search will return everything that contains those two words but not necessarily in the same phrase, thus returning too many GO terms.

Enter your query and click **Find Terms** (**Figure 7A**). When you find the appropriate GO term and definition that describes the molecular function, biological process, or cellular component, click on that term to select it (**Figure 8**). This will direct you to a page showing the detailed information about the term you selected (e.g., parent term and child term relationships or more general terms vs. specific terms) (**Figure 9**). Click **Select this term** if you are satisfied with it (**Figure 9A**); you will be redirected to a page similar to **Figure 6** (see above) with the GO term filled in.

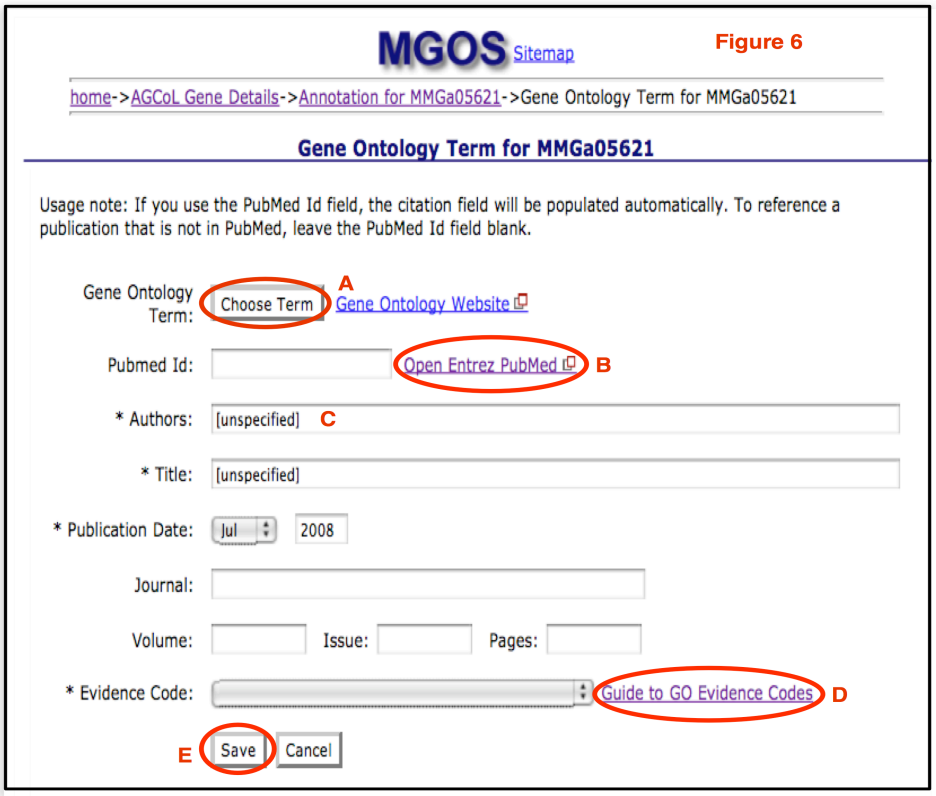
Next, insert the publication that provides the evidence for the GO term. There are two ways to insert your citation:

1) Insert the PubMed ID (PMID) which can be found on either the summary or the abstract page of the article in PubMed. You can use the **Open Entrez PubMed** link to search for the article (**Figure 6B**; this will open a new page).

2) You can manually type (or copy and paste) in the citation if it is not in PubMed (**Figure 6C**; e.g., meeting abstract or dissertation).

If the article is in PubMed, please use the PMID because PubMed and the annotation databases can then be linked by this identifier. After you have inserted the citation, select the **Evidence Code** from the drop-down menu indicating the type of data that supports the GO term. To familiarize yourself with the evidence codes, click on the link **Guide to Gene Ontology Evidence Codes** (**Figure 6D**). When you are finished

Figure 6



providing the evidence code. If the correct citation populates the **Publication** field (Figure 6C), click the **Save** button (**Figure 6E**) and all the information you have

Figure 7

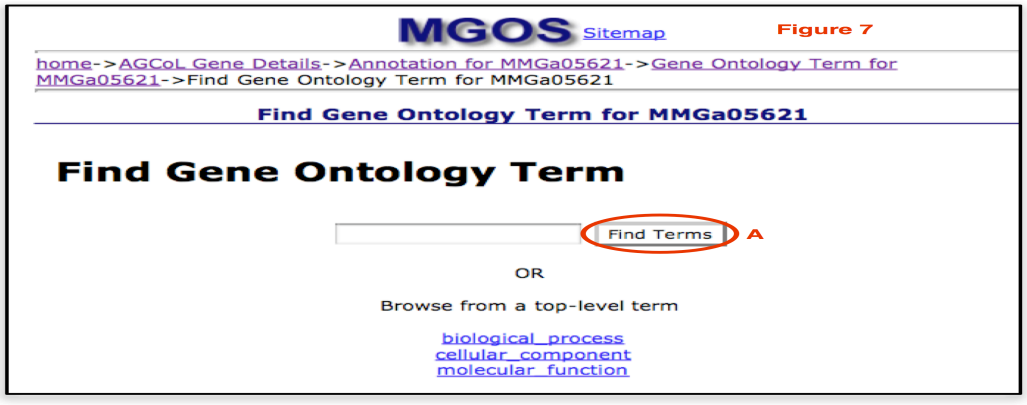
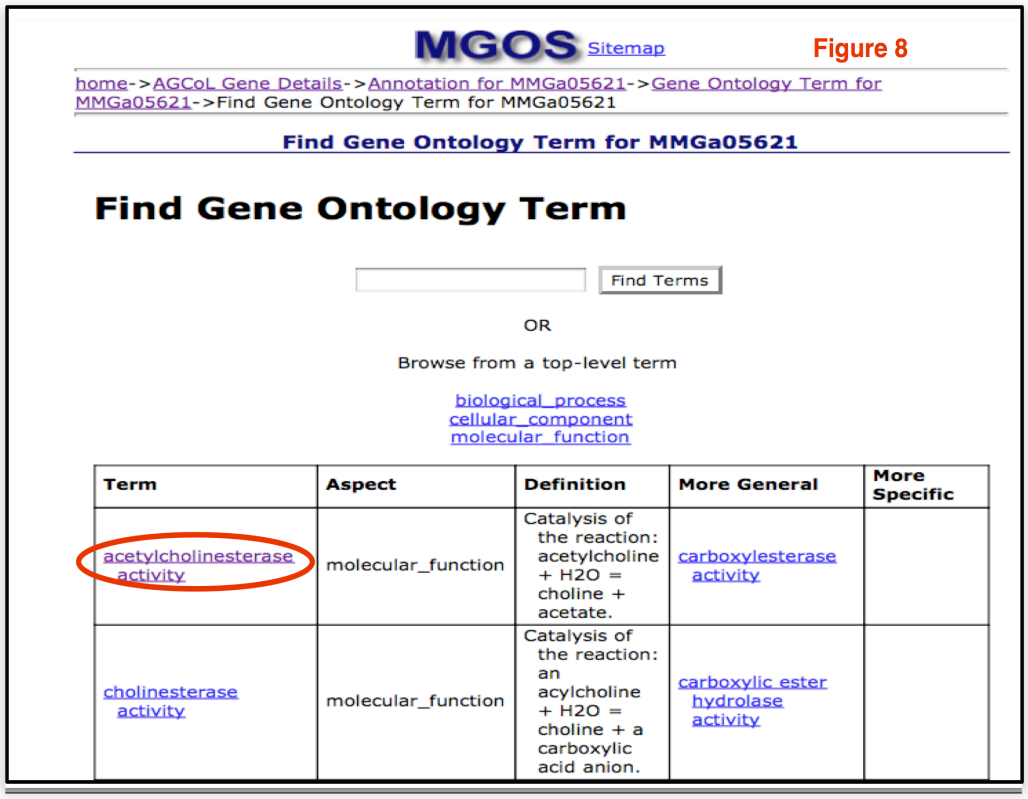


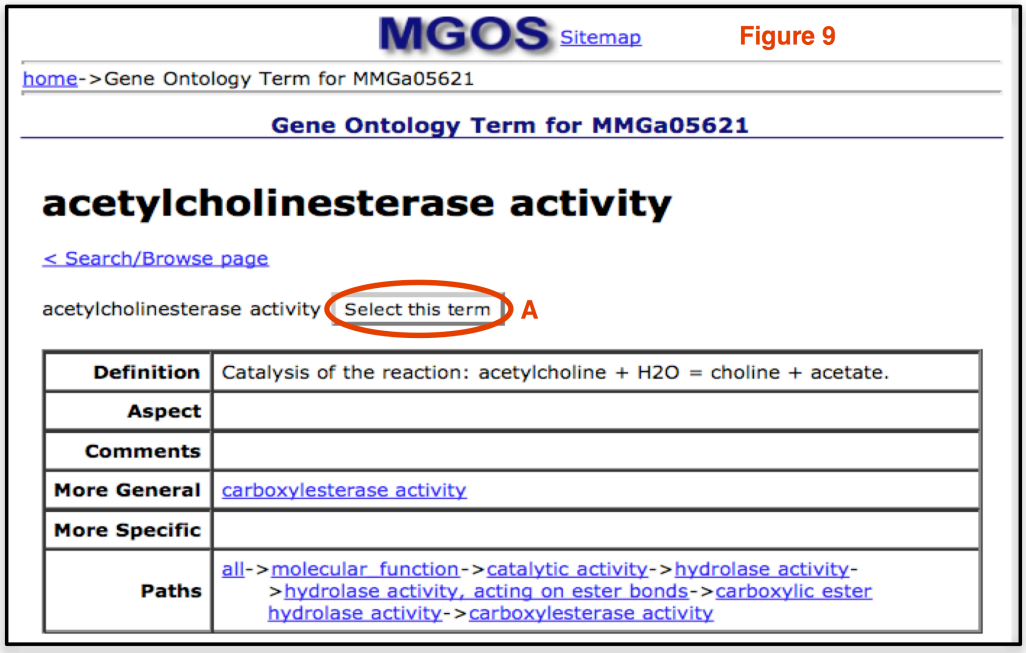
Figure 8



inserted for the GO term will be saved. Only one reference can be cited per GO term. If you have another reference that you wish to cite for the same GO term, you will have to repeat the above process. If you have made a mistake and need to start over, click on the **Cancel** button and none of your Gene Ontology information will be saved. **Please note:** The Gene Ontology is a work in progress. If there are terms that should be defined but are not listed, please contact Anupreet kour (akjosan@email.arizona.edu) or Kevin Greer (kevin@agcol.arizona.edu), they will assist with updating the ontology.

(G) **Add a Fungal Anatomy Term (Figure 4F):** The approach here is similar to the "**Add a GO term"** section. Clicking on the "**Add a Fungal Anatomy Term"** link (**Figure 4F**) will direct you to a search page. Click **Choose Term** in the **Fungal Anatomy Ontology Term** row, enter the anatomy term you are searching for (eg. hyphae) and click **Find Terms**. This will retrieve a page similar to that in **Figure 8** (below). Click on the appropriate term. The next page will display the term, its definition, and the path to the definition (i.e., more general or specific terms). If this term is appropriate, click **Select this term**. You will be redirected to the search page with your term listed in the **Fungal Anatomy Ontology Term** field.
Next add the citation that supports the Fungal Anatomy term either using a PubMed ID or by manually inserting it into the **Publication** field. Select the appropriate **Evidence Code** from the drop-down menu. When you are finished adding the information, click **Save** and your data will be appended to the identifier. The Fungal

Figure 9



Anatomy Ontology is also a work in progress. If there are terms that should be defined but are not listed, please contact Anupreet Kour (akjosan@email.arizona.edu) or Kevin Greer (kevin@agcol.arizona.edu), they will assist with updating the ontology.
If you have made a mistake and need to start over, click on the **Cancel** button. None of your Fungal Anatomy information will be saved in the database.

(H) **Publications:**

(i) **(Figure 4G) Add a citation:** Click on **Add a citation** (**Figure 4G**) to get to the citation page. Either insert the Pub Med ID or manually type (or copy and paste) in the reference information. If you inserted a Pub Med ID, press the enter (or return) key on your keyboard and the citation information will be populated automatically. Click on the "**Save"** button to add the reference to your annotation, or **"Save and New"** if you plan to enter additional references.

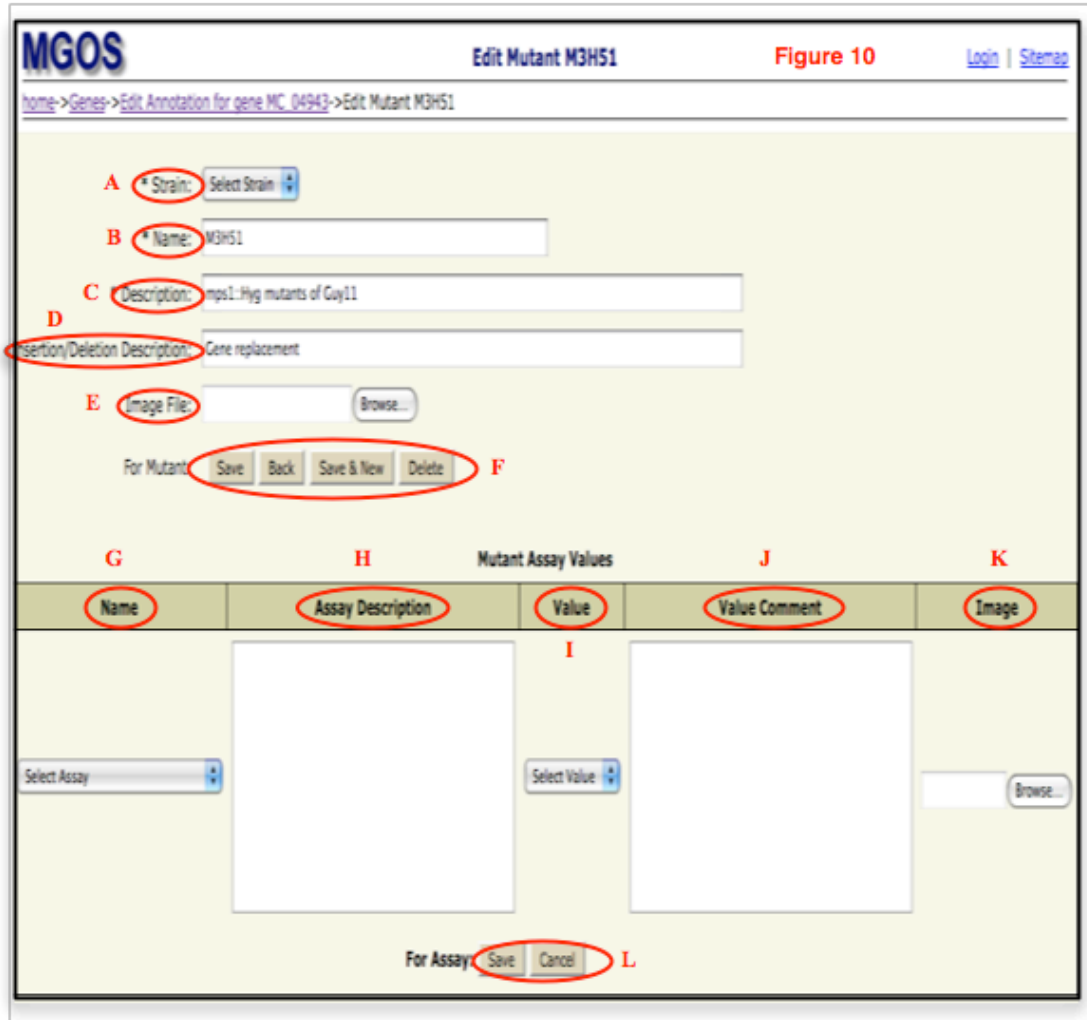
(ii) **(Figure 4H) Edit a citation**: If you want to make changes to a paper that is listed as a citation for the gene, click on the pencil near the citation as shown in **Figure 4H**, and make changes in the page that opens.

(I) **Mutants:**

(i) **(Figure 4*I*) Add a Mutant:** The "**Add a Mutant**" link (**Figure 4*I***) will take you to the page where you can add the name, Description and type of the mutant. Images can be added for the mutant by clicking on Browse in front of Image file field.

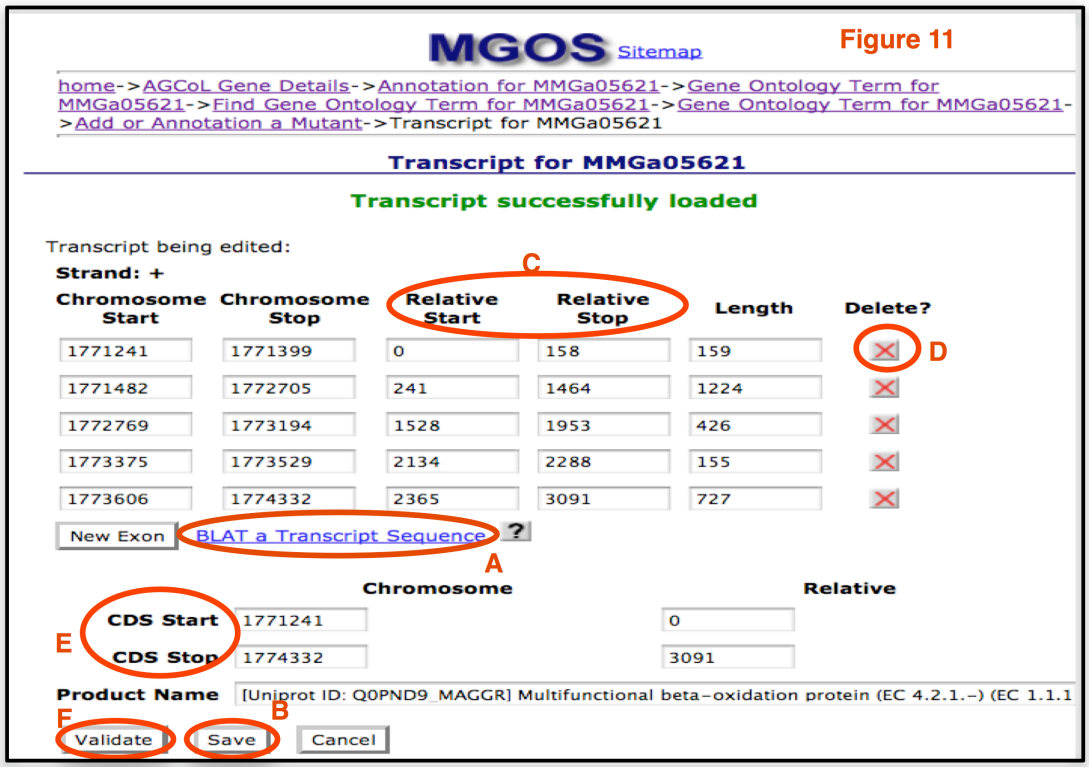
Following options are available on mutant assay page:

Figure 10



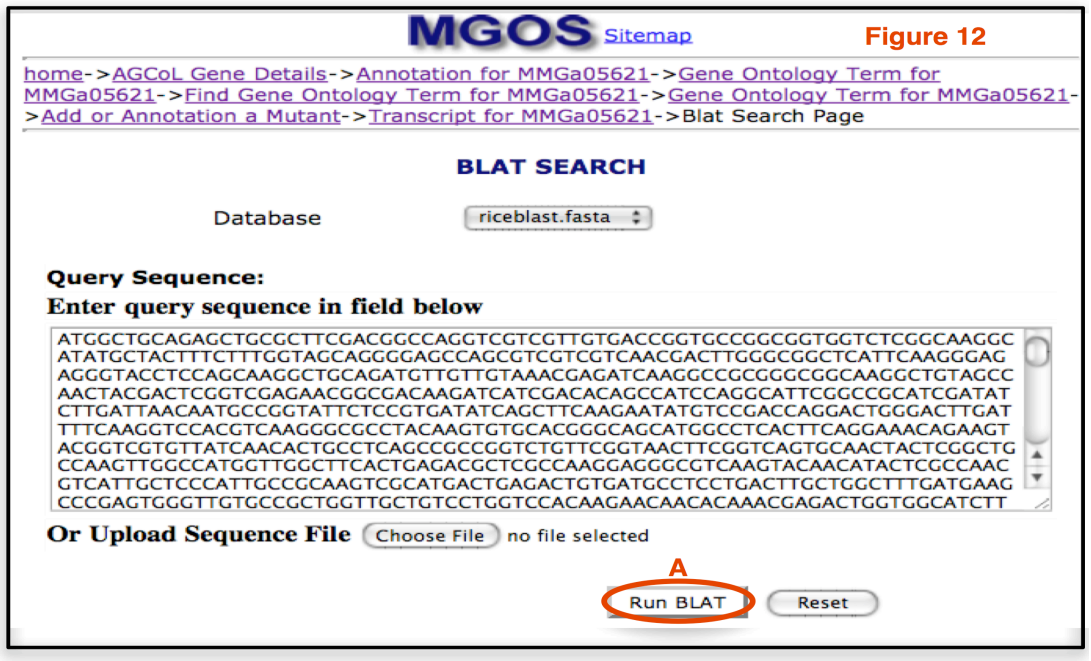
(Figure 10*A*) One can add strain name (eg: 70-15, Guy11 etc.) in this column on which assays had been done.
(Figure 10*B*) Name of the mutant is required for second column.
(Figure 10*C*) Third column is for mutant description i.e how these mutants are generated or what is the resistance cascade introduced in the wild type strain (like hygR).
(Figure 10*D*) Fourth column is for the description of the mutant whether these are insertion/ deletion mutants.
(Figure 10*E*) Under this column one can add some images related to a particular assay.
(Figure 10*F*) The above-entered information can be saved by clicking on save button. Back button directs the link to move to Gene annotation page. Moreover save and new button can be used to save changes in the previous window and open the new window to add new mutant.
(Figure 10*G*) Assay name can be selected by scrolling down select assay option.

Figure 11



(Figure 10*H*) Assay description (like conditions, media, experiment description) can be added in this window.

Figure 12

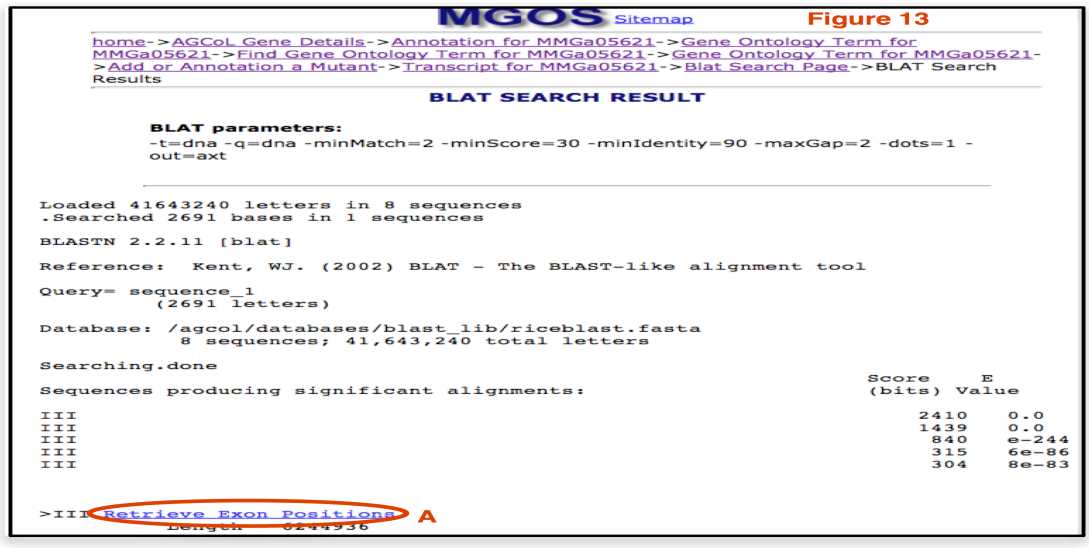


(Figure 10*I*) This scroll down menu contain different values for a particular assay like for appressorium formation it has values like low, normal, aberrant etc. The value can be selected accordingly.
(Figure 10*J*) Value comment window is for the results of a particular assay. One can put different values from the experiments with in a assay.
(Figure 10*K*) Image option here is for adding a picture of the particular assay.
(Figure 10*L*) One can click on save button to save the assay on the window. If you have added something wrong or don't want to add the assay you can click on delete button. It will not save the assays in the window.
After adding the assay one need to click on save button (**Figure 10*F***) to finally save the assay on this link.

(ii) **(Figure 4*J*) Edit Mutants already in the database.** If you want to make changes in already added mutants than you can click on the pencil just near the mutant as shown in **Figure 4*J*** and make changes to the fields described above in **4*I***.

(J) **Transcript annotation:** BLAT option is used to annotate the transcript of a particular locus. Go to the annotation screen for the locus as shown in Figure 4.
(Figure 4K) Click the pencil that is next to the transcript diagram for the locus to get to the transcript maintenance screen. (**Figure 4K** etc.)
(Figure 11) The transcript annotation page will appear.
(Figure 11A) Click the "BLAT a transcript sequence" link (This is next to exon button).
(**Figure 12**) It will open a new window for BLAT
Than paste the new sequence (either EST or Genomic sequence) into the top text box (make sure to overwrite or delete the sequence that is in the box).
(Figure 12A) Click the run BLAT button. Different parameters to run BLAT can be selected on the same page under BLAT options.

Figure 13


MGOS Sitemap
Figure 13
home->AGCoL Gene Details->Annotation for MMGa05621->Gene Ontology Term for MMGa05621->Find Gene Ontology Term for MMGa05621->Gene Ontology Term for MMGa05621->Add or Annotation a Mutant->Transcript for MMGa05621->Blat Search Page->BLAT Search Results
BLAT SEARCH RESULT
BLAT parameters:
-t=dna -q=dna -minMatch=2 -minScore=30 -minIdentity=90 -maxGap=2 -dots=1 -out=axt
Loaded 41643240 letters in 8 sequences
.Searched 2691 bases in 1 sequences
BLASTN 2.2.11 [blat]
Reference: Kent, WJ. (2002) BLAT - The BLAST-like alignment tool
Query= sequence_1
(2691 letters)
Database: /agcol/databases/blast_lib/riceblast.fasta
8 sequences; 41,643,240 total letters
Searching.done
Sequences producing significant alignments:
Score (bits) E Value
III 2410 0.0
III 1439 0.0
III 840 e-244
III 315 6e-86
III 304 8e-83
>III Retrieve Exon Positions
A

(Figure 13) BLAT result screen will appear.
(Figure 13A) Click "Retrieve exon positions". This will link to copy the new coordinates for the exons back into the transcript maintenance screen. (**Figure 11**). The alignment results of the query sequence and the genomic sequence will appear on the same page. Close the BLAT result screen. The transcript maintenance screen should now have the correct exons and corresponding translated sequence (Displayed at the bottom of screen).
(Figure 11B) Than click save button to save it on the screen.
If the desired gene model is not achieved one can go for following steps:
(Figure 11C) If there is problem in the number of exons that are required to be in the sequence that can be set by putting values of nucleotide sequence number in the boxes (namely: Relative start and relative stop) provided on the top of the transcript-editing page.
(Figure 11D) Only required boxes should be kept. The delete button can delete as if some sequence is having two exons than except these two exon boxes all the other.
(Figure 11E) To set transcript coding sequence the relative values of start and stop codon can be put into the CDS start and CDS stop boxes available.
(Figure 11F) After putting these values click on validate button. The final amino acid sequence will appear at the bottom of the page.
(Figure 11B) These changes can be saved by clicking on save button.